PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61K 38/43, 38/48, G01N 33/48, C12Q 1/37 — A1

(11) International Publication Number: WO 95/22985

(43) International Publication Date: 31 August 1995 (31.08.95)

(21) International Application Number: PCT/IT95/00018

(22) International Filing Date: 14 February 1995 (14.02.95)

(30) Priority Data:
RM94A000092 23 February 1994 (23.02.94) IT

(71) Applicant *(for all designated States except US)*: ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI S.P.A. [IT/IT]; Via Pontina Km. 30.600, I-00040 Pomezia (IT).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: DE FRANCESCO, Raffaele [IT/IT]; Via Devich, 46, I-00146 Rome (IT). FAILLA, Cristina [IT/IT]; Viale Libia, 121, I-00199 Rome (IT). TOMEI, Licia [IT/IT]; Viale Augusto, 122, I-80123 Napoli (IT).

(74) Agents: DI CERBO, Mario et al.; Società Italiana Brevetti S.p.A., Piazza di Pietra, 39, I-00187 Roma (IT).

(81) Designated States: AU, BR, CA, CN, JP, RU, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR REPRODUCING IN VITRO THE PROTEOLYTIC ACTIVITY OF THE NS3 PROTEASE OF HEPATITIS C VIRUS (HCV)

(57) Abstract

This is a method for reproducing in vitro the serine protease activity associated with the HCV NS3 protein, that comprises the use both of sequences contained in NS3 and sequences contained in NS4A. This method takes advantage of the ability of the HCV NS4A protein, or sequences contained therein, to act as a cofactor of the serine protease activity or more generally of the enzymatic activities associated with NS3. Optimal serine protease activity is obtained when NS4A is present in a molar ratio of at least 1:1 with NS3. NS3 and NS4A can also be incorporated in the reaction mixture as NS3-NS4A precursor, as this precursor will generate, by means of an autoproteolytic event, equimolar amounts of NS3 and NS4A. It is also possible to mutate the cleavage site between NS3 and NS4A in a precursor, so that NS4A remains covalently bonded to NS3. The sequences that do not influence the proteolytic activity of NS3 can subsequently be removed from said non-proteolyzable precursor. The invention also relates to a composition of matter that comprises sequences contained in NS3 and NS4A, and to the use of these compositions for the setup of an enzymatic test capable of selecting, for therapeutic purposes, compounds that inhibit the enzymatic activity associated with NS3. The figure shows plasmidic vectors used in the method to activate HCV NS3 protease in cultivated cells and in vitro.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

METHOD FOR REPRODUCING IN VITRO THE PROTEOLYTIC ACTIVITY OF THE NS3 PROTEASE OF HEPATITIS C VIRUS (HCV)

DESCRIPTION

The present invention has as its subject a method for reconstituting the serine protease activity associated with the HCV NS3 protein, which makes use of the ability of the HCV protein NS4A, or sequences contained therein, to act as a cofactor of the serine protease activity or more generally speaking of enzymatic activities associated with NS3.

As is known, the hepatitis C virus (HCV) is the main etiological agent of non-A, non-B hepatitis (NANB). It is estimated that HCV causes at least 90% of post-transfusional NANB viral hepatitis and 50% of sporadic NANB hepatitis. Although great progress has been made in the selection of blood donors and in the immunological characterization of blood used for transfusions, there is still a high level of acute HCV infection among those receiving blood transfusions (one million or more infections every year throughout the world). Approximately 50% of HCV-infected individuals develop cirrhosis of the liver within a period that can range from 5 to 40 years. Furthermore, recent clinical studies suggest that there is a correlation between chronic HCV infection and the development of hepatocellular carcinoma.

HCV is an enveloped virus containing an RNA positive genome of approximately 9.4 kb. This virus is a member of the Flaviviridae family, the other members of which are the flaviviruses and the pestiviruses. The RNA genome of HCV has recently been mapped. Comparison of sequences from the HCV genomes isolated in various parts of the world has shown that these sequences can be extremely heterogeneous. The majority of the HCV genome is occupied by an open reading frame (ORF) that can vary between 9030 and 9099 nucleotides. This ORF codes for a single viral polyprotein, the length of which can vary

from 3010 to 3033 amino acids. During the viral infection cycle, the polyprotein is proteolytically processed into the individual gene products necessary for replication of the virus. The genes coding for HCV structural proteins are located at the 5'-end of the ORF, whereas the region coding for the non-structural proteins occupies the rest of the ORF.

The structural proteins consist of C (core, 21 kDa), E1 (envelope, gp37) and E2 (NS1, gp61). C is a non-glycosylated protein of 21 kDa which probably forms the viral nucleocapsid. The protein E1 is a glycoprotein of approximately 37 kDa and it is believed to be a structural protein for the outer viral envelope. E2, another membrane glycoprotein of 61 kDa, is probably a second structural protein in the outer envelope of the virus.

The non-structural region starts with NS2 (p24), a hydrophobic protein of 24 kDa whose function is unknown. NS3, a protein of 68 kDa which follows NS2 in the polyprotein, is predicted to have two functional domains: a serine protease domain in the first 200 amino-terminal amino acids, and an RNA-dependent ATPase domain at the carboxy terminus. The gene region corresponding to NS4 codes for NS4A (p6) and NS4B (p26), two hydrophobic proteins of 6 and 26 kDa, respectively, whose functions have not yet been clarified. The gene corresponding to NS5 also codes for two proteins, NS5A (p56) and NS5B (p65), of 56 and 65 kDa, respectively. An amino acid sequence present in all the RNA-dependent RNA polymerases can be recognized within the NS5 region. This suggests that the NS5 region contains parts of the viral replication machinery.

Various molecular biological studies indicate that the signal peptidase, a protease associated with the endoplasmic reticulum of the host cell, is responsible for proteolytic processing in the non-structural region, that is to say at sites C/E1, E1/E2 and E2/NS2. The

serine protease contained in NS3 is responsible for cleavage at the junctions between NS3 and NS4A, between NS4A and NS4B, between NS4B and NS5A and between NS5A and NS5B. In particular it has been found that the cleavage performed by this serine protease leaves a residue of cysteine or threonine on the amino-terminal side (position P1) and an alanine or serine residue on the carboxy-terminal side (position P1') of the scissile bond. A second protease activity of HCV appears to be responsible for the cleavage between NS2 and NS3. This protease activity is contained in a region comprising both part of NS2 and the part of NS3 containing the serine protease domain, but does not use the same catalytic mechanism.

In the light of the above description, the NS3 protease is considered a potential target for the development of anti-HCV therapeutic agents. However, the search for such agents has been hampered by the evidence that the serine protease activity displayed by NS3 in vitro is too low to allow screening of inhibitors.

It has now been unexpectedly found that this important limitation can be overcome by adopting the method according to the present invention, which also gives additional advantages that will be evident from the following.

According to the present invention, the method to reproduce in vitro the proteolytic activity of the protease NS3 of HCV is characterized by using in the reaction mixture, both sequences contained in NS3 and sequences contained in NS4A.

Optimal serine protease activity is obtained when NS4A is present in a ratio of 1:1 with NS3.

NS3 and NS4A can also be incorporated in the reaction mixture as NS3-NS4A precursor, as this precursor will generate, by means of an autoproteolytic event, equimolar amounts of NS3 and NS4A.

It is also possible to mutate the site of cleavage between NS3 and NS4A, in a precursor, so that NS4A remains covalently bound to NS3. The sequences that do not influence the proteolytic activity of NS3 can subsequently be removed from this non-proteolyzable precursor.

The invention also extends to a new composition of matter, characterized in that it comprises proteins whose sequences are described in SEQ ID NO:1 and SEQ ID NO:2 or sequences contained therein or derived therefrom. It is understood that these sequences may vary in different HCV isolates, as all the RNA viruses show a high degree of variability. This new composition of matter has the proteolytic activity necessary to obtain the proteolytic maturation of several of the non-structural HCV proteins.

The present invention also has as its subject the use of these compositions of matter in order to prepare an enzymatic assay capable of identifying, for therapeutic purposes, compounds that inhibit the enzymatic activity associated with NS3, including inhibitors of the interaction between NS3 and NS4A.

Up to this point a general description has been given of the present invention. With the aid of the following examples, a more detailed description of specific embodiments thereof will now be given, in order to give a clearer understanding of its objects, characteristics, advantages and method of operation.

The figure illustrates plasmid vectors used in the method to activate the HCV NS3 protease in cultivated cells and in vitro (example 1 and example 2).

EXAMPLE 1

METHOD OF ACTIVATION OF HCV NS3 SERINE PROTEASE IN CULTIVATED CELLS

Plasmid vectors were constructed for expression of NS3, NS4A and other non-structural HCV proteins in HeLa cells. The plasmids constructed are schematically

illustrated in figure 1. Selected fragments of the cDNA corresponding to the genome of the HCV BK isolate (HCV-BK) were cloned downstream of the promoter of the bacteriophage T7 in the plasmid vector pCite-1^R (Novagen). This expression vector contains the internal ribosome entry site of the encephalomyocarditis virus, so as to guarantee an effective translation of the messenger RNA transcribed from promoter T7, even in the absence of a CAP structure.

The various fragments of HCV-BK cDNA were cloned into the plasmid pCite-1^R using methods known in molecular biology practice. pCite(NS3) contains the portion of the HCV-BK genome comprised between nucleotides 3351 and 5175 (amino acids 1007-1615 of the polyprotein). pCite(NS4B/5A) contains the portion of the HCV-BK genome comprised between the nucleotides 5652 and 7467 (amino acids 1774-2380). pCite(NS3/4A) contains the portion of the HCV-BK genome comprised between the nucleotides 3711 and 5465 (amino acids 991 and 1711). pCite(NS4A) contains the portion of the HCV-BK genome comprised between the nucleotides 5281 and 5465 (amino acids 1649-1711). pCite(NS5AB) contains the portion of the HCV-BK genome comprised between the nucleotides 6224 and 9400 (amino acids 1965-3010). The numbering given above agrees with the sequences for the genome and the polyprotein given for HCV-BK in Takamizawa et al, Structure and organization of the hepatitis C virus genome isolated from human carriers, (1991), J. Virol. 65, 1105-1113.

In order to obtain efficient expression of the various portions of the HCV polyprotein, the HeLa cells were infected with vTF7-3, a recombinant vaccinia virus which allows synthesis of the RNA polymerase of the bacteriophage T7 in the cytoplasm of infected cells. These cells, after infection, were then transfected with plasmid vectors selected from among those described in figure. The HeLa cells thus infected and transfected

were then metabolically labelled with [^{35}S]methionine and the recombinant proteins encoded by the various plasmids could be identified by immunoprecipitation with polyclonal rabbit antibodies that recognize NS3, NS4 or NS5A. The method described in the present example for analysis of recombinant HCV proteins has already been described in L. Tomei et al, "NS3 is a serine protease required for processing of hepatitis C virus polyprotein", J. Virol. (1993) 67, 1017-1026 and in the bibliography mentioned therein.

By transfecting the plasmid pCite(NS3) into the HeLa cells infected with vTF7-3, it is possible to observe the synthesis of a protein containing the catalytic domain of the HCV NS3 protease. pCite(NS4B5A) codes for a portion of the HCV polyprotein containing a peptide bond, at the junction between NS4B and NS5B, which would be expected to be hydrolyzed by the serine protease activity associated with NS3. However, when pCiteNS3 is cotransfected with pCiteNS4B5A, there is no evidence of proteolytic celavage. Conversely, when the NS3 serine protease domain is expressed in combination with NS4A the proteolytic cleavage of the precursor encoded by pCite(NS4B5A) can take place normally. Coexpression of the NS3 serine protease domain and 4A can be achieved, for example, by transfection with equimolar amounts of the plasmids pCite(NS3) and pCite(NS4A), by transfection of a plasmid coding for a precursor containing both NS3 and NS4A [pCite(NS34A)], or by transfection of a derivative of the latter plasmid to which all the sequence that are not relevant for proteolysis have been deleted [pCite(NS34A)], or by transfection of a derivative of the latter plasmid to which all the sequence that are not relevant for proteolysis have been deleted [pCite(NS3Δint1237-1635)]. NS4A expressed transiently in HeLa cells can thus activate the proteolytic activity associated with NS3, which otherwise would not be seen.

EXAMPLE 2

METHOD FOR ACTIVATION OF THE HCV SERINE NS3 PROTEASE IN AN IN VITRO TRANSLATION ASSAY

The plasmids described in figure 1 can also be used for in vitro synthesis of mRNA coding for the respective HCV proteins using the purified RNA polymerase enzyme of the phage T7 (Promega).

Generally the plasmids derived from pCite-1^R were linearized using suitable restriction enzymes and transcribed using the protocols supplied by the manufacturer (Promega). These synthetic mRNA, could later be used to synthesize the corresponding proteins in extracts of rabbit reticulocytes in the presence of canine pancreas microsomal membranes. The reticulocyte extracts, the canine pancreas microsomal membranes, like all the other material required, were purchased from Promega, which also supplied the instructions for the in vitro protein syntheses process described above.

Programming the in vitro translation mixture with mRNA transcribed from pCite(NS3) it is possible to observe synthesis of a protein with the expected molecular weight (68 kDa) containing the entire NS3 serine protease domain. The mRNA transcribed from pCite(NS5AB) guides the synthesis of a precursor of 115 kDa which contains NS5A and NS5B and is thus a substrate for the proteolytic activity associated with NS3.

However, when the two proteins, containing the NS3 serine protease domain and the substrate with the site corresponding to the junction between NS5A and NS5B, are synthesized in the same reaction mixture, there is no clear evidence of the proteolytic activity of NS3.

On the contrary, the mRNA transcribed from pCite(NS34A) is translated into a precursor protein of approximately 76 kDa which self-processes proteolytically in vitro to give equimolar amounts of two proteins of 70 kDa and 6kDa, containing NS3 and NS4A, respectively.

If, in addition to the mRNA transcribed from pCite(NS34A), the mRNA transcribed from pCite(NS5AB) is included in the in vitro translation mixture, there can be observed, in addition to the self-proteolysis at the site between NS3 and NS4A, the generation of two new proteins of 56 kDA and 65 kDA which contain NS5A and NS5B, respectively. These proteins represent the product of proteolysis of the precursor containing NS5A and NS5B by NS3. Similarly, the 56 kDa and 65 kDa protein products, generated proteolytically from the NS5AB precursor, are obtained if the mRNA transcribed from pCite(NS3Δint1237-1635) is cotranslated with the mRNA translated from pCite(NS5AB).

This result can be summarized by stating that, in vitro, the protease domain of NS3 alone is not capable of exhibiting protease activity on a substrate containing NS5A and NS5B. However, the serine protease activity of NS3 becomes evident if another protein sequence containing NS4A is present in addition to the NS3 protease domain.

EXAMPLE 3

METHOD OF ACTIVATION OF THE HCV NS3 PROTEASE USING A SYNTHETIC PEPTIDE CONTAINING NS4A SEQUENCES

A synthetic peptide containing the sequence SEQ ID NO:3 was synthesized on solid phase. This sequence is derived from the C-terminal portion of SEQ ID NO:2. Synthesis of the peptide took place on solid phase according to processes known to those operating in this field. In this peptide, the carboxy terminal cysteine has been replaced with alpha-aminobutyric acid (Abu).

This peptide was added to an in vitro translation mixture simultaneosly programmed with the mRNAs transcribed from the plasmids pCite(NS3) and pCite(NS5AB).

It was thus possible to observe the proteolytic activity associated with the serine protease domain of NS3, which resulted in the proteolytic cleavage of the

substrate in the two products containing the proteins NS5A and NS5B. This activity is dependent on the simultaneous presence of the NS3 serine protease domain and the synthetic peptide with the sequence SEQ ID NO:3.

EXAMPLE 4

METHOD OF ASSAY OF A RECOMBINANT HCV NS3 SERINE PROTEASE ON A PEPTIDE SUBSTRATE

The plasmid pT7-7 NS3(1027-1206), described in figure 1 and in Example 4, was constructed in order to allow expression in E. coli of the protein fragment comprised between amino acid 1 and amino acid 180 of Seq.ID NO 1. Such fragment contains the serine protease domain of NS3, as determined experimentally. The fragment of HCV cDNA coding for NS3 fragment just described was cloned in the pT7-7 plasmid, an expression vector that contains the T7 RNA polymerase promoter ϕ 10 and the translation start site for the T7 gene 10 protein (Studier and Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, (1986), J. Mol. Biol. 189, p. 113-130). The cDNA fragment coding for the NS3 serine protease domain as defined above was cloned downstream of the bacteriophage T7 promoter and in frame with the first ATG codon of the T7 gene 10 protein, using methods that are known to the molecular biology practice. The pT7-7 plasmid also contains the gene for the ß-lactamase enzyme, which can be used as a marker of selection of E. coli cells transformed with plasmids derived with pT7-7.

The plasmid pT7-7 NS3(1027-1206) is then transformed in the E. coli strain BL21(DE53), which is normally employed for high-level expression of genes cloned into expression vectors containing T7 promoter. In this strain of E. coli, the T7 gene polymerase is carried on the bacteriophage λDE53, which is integrated into the chromosome of BL21. Expression from the gene of interest is induced by addition of isopropylthiogalactoside (IPTG) to the growth medium

according to a procedure that has been previously described (Studier and Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, (1986), J. Mol. Biol- 189, p. 113-130).

The recombinant NS3 fragment containing the serine protease domain could be purified from E. coli BL21(DE53) transformed with the plasmid pT7-7 NS3(1027-1206) by the procedure summarized below.

In brief, E. coli BL21(DE53) cells harboring the pT7-7 NS3(1027-1206) plasmid were grown at 37°C to an optical density at 600 nm of around 0.8 absorbance units. Thereafter, the medium was cooled down to 22°C and production of the desired protein induced by addition of IPTG to a final concentration of 0.4 mM. After 4-6 hours at 22°C in the presence of IPTG, cells were harvested and lysed by means of a French-pressure cell in a buffer containing 20 mM sodium phosphate pH 6.5, 0.5% (w/v) (3-[(3-cholamidopropyl)-dimethylammonio] 1-propanesulfonate (CHAPS), 50% (v/v) glycerol, 10 mM dithiothreitol and 1 mM EDTA (lysis buffer). The cell debris was removed by centrifugation (1 hour at 120000 x g) and the resulting pellet resuspended in lysis buffer, digested with DNAse I, re-homogenized and re-centrofuged as described above. S-Sepharose Fast Flow ion exchange resin (Pharmacia) pre-equilibrated in lysis buffer was added to the pooled supernatants (30% v/v) and the slurry was stirred for 1 hour at 4°C. The resin was sedimented and washed extensively with lysis buffer and poured into a chromatography column. The NS3 protease was eluted from the resin by applying a 0-1 M NaCl gradient. The protease-containing fractions equilibrated with 50 mM sodium phosphate buffer pH 7.5, 10% (v/v) glycerol, 0.5% (w/v) CHAPS and 2 mM dithiothreitol. The protein was 90-95% pure after this step. Purification to >98% was achieved by subsequent chromatography on Heparin Sepharose equilibrated with 50 mM Tris pH 7.5, 10% (v/v) glycerol, 0.5% (w/v) CHAPS and 2 mM

dithiothreitol. Elution of the NS3 protease from this column was achieved by applying a linear 0-1 M NaCl gradient.

The concentration of the purified protein was determined by the Bio-Rad protein assay (Bio-Rad cat. 500-0006).

The recombinant NS3 serine protease produced according to the above procedure in E. coli could be assayed for activity by cleaving a substrate that provides detectable cleavage producs. The signal is preferably detectable by colorimetric or fluorometric means. Methods such as HPLC and the like are also suitable.

For example, we used, as a substrate, synthetic peptides corresponding to the NS4A/4B junction of the HCV polyprotein and containing the aminoacid sequence SEQ ID NO:4 or part of it.

Alternatively, peptide esters, having the general structure indicated in SEQ ID NO:5.

The activity assay is performed by incubating 5-1000 μM substrate and 0.05-1 μM protease in buffer containing 25 mM Tris/HCl pH 7.5, 3 mM dithiothreitol, 0.5% (w/v) CHAPS and 10% (v/v) glycerol for 1-3 hours at 22°C. The reaction is stopped by addition of trifluoracetic acid to yield a final concentration of 0.1% (w/v).

The reaction products are then separated by HPLC on a C18 reverse phase column and quantitated according to their absorbance of the far UV light.

The proteolytic activity displayed by recombinant NS3 serine protease purified from E. coli is very low when the activity assay is performed as described above. However, we found that increasing amounts of the synthetic peptide described in SEQ ID NO:3 stimulate the proleolytic activity of the recombinant NS3 serine protease up to 20-fold. Maximal activity is reached when

the recombinant NS3 serine protease and the synthetic peptide are present in equimolar amounts.

The assay described above can be used for the search of protease inhibitors. Because the activity of NS3 protease in such assay depends on the interaction of the NS3 serine protease domain with amino acid sequences derived from NS4A, it is also possible, by using the assay described above, to search for antagonists of the interaction between NS3 and NS4A that will ultimately inhibit the proteolytic activity associated with NS3.

EXAMPLE 5

DETAILED CONSTRUCTION OF THE PLASMIDS IN THE SOLE FIGURE

pCite(NS3) contains the portion of the HCV-BK genome comprised between nucleotides 3351 and 5175 (amino acids 1007-1615 of the polyprotein). Construction of this plasmid has been described in L. Tomei et al, "NS3 is a serine protease required for processing of hepatitis C virus polyprotein", J. Virol (1993) 67, 1017-1026.

pCite(NS4B/5A) was obtained by cloning a ScaI-BamHI fragment derived from the plasmid pCite(NS4-5), described in Tomei et al, into pCite(NS3) that was previously digested with MscI and BamHI. pCite(NS4B/5A) contains the portion of the HCV genome comprised between nucleotides 5652 and 7467 (amino acids 1774-2380 of the polyprotein).

pCite(NS5AB) codes for a protein that comprises the sequence from amino acid 1965 to amino acid 3010 of the HCV-BK polyprotein. To construct this plasmid, the plasmid pCite(SX) described in Tomei et al (1993), supra, was first digested with AseI and treated with the Klenow fragment of the DNA polymerase. After inactivation of the Klenow enzyme, the plasmid was digested with XbaI. The resulting cDNA fragment, containing the region between nucleotides 6224 and 9400, was purified and inserted into the BstXI and XbaI sites

of the vector pCite-1^R, after blunting the end generated by BstXI with T4 DNA polymerase.

pCite(NS3/4A) was obtained as follows. A cDNA fragment, corresponding to the region between nucleotides 3711 and 5465 of the HCV-BK genome, was synthesized by means of polymerase chain reaction (PCR) using sequence-specific oligonucleotides as primers. A UAG stop codon was suitably included in the antisense oligonucleotide. After PCR amplification, the resulting cDNA was cleaved at the 5' end with SAlI and the product of 750 pairs of bases cloned directionally into the SalI and NheI sites of the plasmid pCite(SX), after blunt-ending the NheI end with the Klenow fragment of the DNA polymerase. The resulting plasmid codes for the portion of HCV-BK polyprotein comprised between amino acids 991 and 1711.

For the construction of pCite(NS4A), a cDNA fragment, corresponding to the region between the nucleotides 5281 and 5465 of the HCV-BK genome (amino acids 1649-1711), was obtained by polymerase chain reaction (PCR) amplification with sequence-specific oligonucleotides as primers. The cDNA resulting from the PCR amplification was subsequently cloned into the BstxI and StuI sites of the plasmid pCite-1^R, after blunt-ending the BstXI digested end with the DNA polymerase of the bacteriophage T4.

pCite(NS3Δint1237-1635) is a derivative of pCite(NS3/4A) from which all the sequences comprised between nucleotide 4043 and nucleotide 5235 have been deleted. It was obtained by digesting pCite(NS3/4A) with BsteII and partially with ScaI. The fragment containing the deletion of interest was then circularised by use of T4 DNA ligase. This plasmid codes for a protein that has the same amino- and carboxy-terminal ends as that encoded by pCite(NS3/4A), but all the amino acid residues comprised between amino acid 1237 and amino acid 1635, experimentally found to be

dispensible for the serine-protease NS3 activity, have been deleted.

pT7-7[NS3(1027-1206)] contains the HCV sequence from nucleotide 3411 to nucleotide 3951, encoding the HCV NS3 fragment comprised between amino acid 1027 and amino acid 1206. In order to obtain this plasmid, a DNA fragment was generated by amplification of HCV cDNA by the polymerase chain reaction (PCR) using the oligo nucleotides referred to as SEQ ID NO:6 and SEQ ID NO:7. The cDNA fragment obtained by PCR was phosphorylated, digested with Nde I and subsequently cloned downstream of the bacteriophage T7 promoter, following immediately the first ATG codon of the T7 gene 10 protein in the vector pT7-7 previously digested with Nde I and Sma I (Studier and Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, (1986), J. Mol. Biol. 189, p. 113-130). It is to note that an amber codon was inserted immediately following the HCV-derived sequence.

SEQUENCE LISTING

GENERAL INFORMATION

(i) APPLICANT: ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI S.p.A.

(ii) TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE PROTEOLYTIC ACTIVITY OF THE NS3 PROTEASE OF HEPATITIS C VIRUS (HCV)

(iii) NUMBER OF SEQUENCES: 7

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Società Italiana Brevetti

(B) STREET: Piazza di Pietra, 39

(C) CITY: Rome

(D) COUNTRY: Italy

(E) POSTAL CODE: I-00186

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk 3.5" 1.44 MBYTES

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS Rev. 5.0

(D) SOFTWARE: Microsoft Wordstar 4.0

(viii) ATTORNEY INFORMATION

(A) NAME: DI CERBO, Mario (Dr.)

(C) REFERENCE: RM/X88350/PCT-DC

(ix) TELECOMMUNICATION INFORMATION

(A) TELEPHONE: 06/6785941

(B) TELEFAX: 06/6794692

(C) TELEX: 612287 ROPAT

(1) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 631 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: No

(iv) ANTISENSE: No

(v) FRAGMENT TYPE: internal fragment

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hepatitis C Virus
(C) ISOLATE : BK
(vii) IMMEDIATE SOURCE: cDNA clone pCD(38-9.4) described by Tomei et al. in 1993
(ix) FEATURE:
(A) NAME: NS3 Serine Protease Domain
(B) LOCATION: 1-180
(C) IDENTIFICATION METHOD: Experimentally
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys
1               5                  10                  15
Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu
            20                  25                  30
Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val
        35                  40                  45
Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu
    50                  55                  60
Ala Ala Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Cal Asp Gln
65                  70                  75                  80
Asp Leu Val Gly Trp Pro Lys Pro Pro Gly Ala Arg Ser Leu Thr Pro
                85                  90                  95
Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp
            100                 105                 110
Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser
        115                 120                 125
Pro Arg Pro Cal Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu
    130                 135                 140
Cys Pro Phe Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr
145                 150                 155                 160
Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu
                165                 170                 175
Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala
            180                 185                 190
Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser
        195                 200                 205
Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys
```

```
    210                 215                 220
Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala
225                 230                 235                 240
Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val
                245                 250                 255
Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys
            260                 265                 270
Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile
        275                 280                 285
Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly
    290                 295                 300
Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu
305                 310                 315                 320
Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile
                325                 330                 335
Glu Lgu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys
            340                 345                 350
Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys
        355                 360                 365
His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu
    370                 375                 380
Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile
385                 390                 395                 400
Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
                405                 410                 415
Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val
            420                 425                 430
Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr
        435                 440                 445
Thr Thr Val Pro Gln Aps Ala Val Ser Arg Ser Gln Arg Arg Gly Arg
    450                 455                 460
Thr Gly Arg Gly Arg Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu
465                 470                 475                 480
Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp
                485                 490                 495
Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg
            500                 505                 510
```

```
Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His
        515                 520                 525
Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala
    530                 535                 540
His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu
545                 550                 555                 560
Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro
                565                 570                 575
Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu
            580                 585                 590
His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu
        595                 600                 605
Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser
    610                 615                 620
Ala Asp Leu Glu Val Val Thr
625                 630
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS
- (A) LENGTH: 54 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: polypeptide

(iii) HYPOTHETICAL: No

(iv) ANTISENSE: No

(v) FRAGMENT TYPE: Internal

(vii) IMMEDIATE SOURCE: cDNA Clone (SEE SEQ ID NO:1)

(ix) FEATURE:
- (A) NAME: NS4A Protein
- (C) IDENTIFICATION METHOD: Experimentally

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr
1               5                   10                  15
Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser
            20                  25                  30
Gly Arg Pro Ala Ile Val Pro Asp Arg Glu Leu Leu Tyr Gln Glu Phe
        35                  40                  45
```

Asp Glu Met Glu Glu Cys
50

(3) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: polypeptide

(iii) HYPOTHETICAL: No

(iv) ANTISENSE: No

(v) FRAGMENT TYPE: internal

(vii) IMMEDIATE SOURCE:

(A) SYNTHESIS: Solid phase peptide synthesis

(ix) FEATURE:

(A) NAME: Cofactor of NS3 serine protease

(C) IDENTIFICATION METHOD: experimentally

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala
1 5 10 15

Ile Val Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu Met Glu
20 25 30

Glu Abu

CLAIMS

1. A method for reproducing in vitro the proteolytic activity of the HCV NS3 protease, characterized in that both sequences contained in NS3 and sequences contained in NS4A are used in the reaction mixture.

2. The method for reproducing in vitro the proteolytic activity of the HCV NS3 protease according to claim 1, in which NS4A is present in a ratio of 1:1 with respect to NS3.

3. The method for reproducing in vitro the proteolytic activity of the HCV NS3 protease according to claim 1 or 2, in which NS3 and NS4A are incorporated in the reaction mixture as NS3-NS4A precursor, said precursor generating, by means of an autoproteolytic event, equimolar amounts of NS3 and NS4A.

4. The method for reproducing in vitro the proteolytic activity of the HCV NS3 protease according to any one of the preceding claims, in which the cleavage site between NS3 and NS4A is mutated in a precursor, so that NS4A remains covalently bonded to NS3, it being subsequently possible to remove from said non proteolyzable precursor the sequences that do not influence the proteolytic activity of NS3.

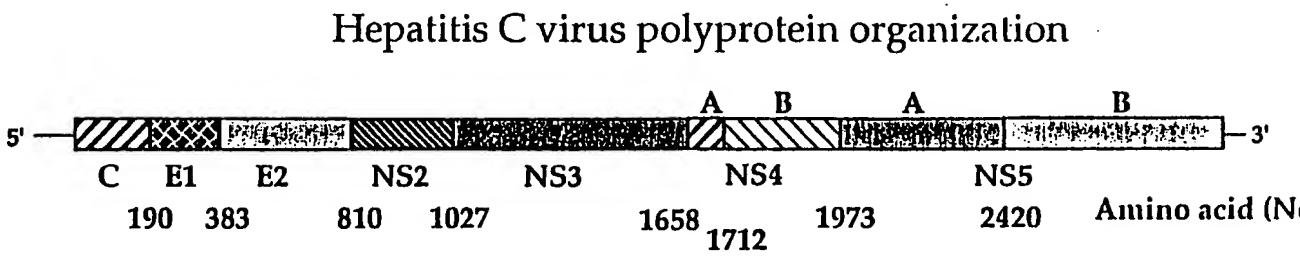
5. A composition of matter, characterized in that it contains NS3 and NS4A sequences according to claims 1 to 4.

6. The composition of matter according to claim 5, comprising the proteins whose sequences are described in SEQ ID NO:1 and SEQ ID NO:2 or sequences contained therein or derived therefrom.

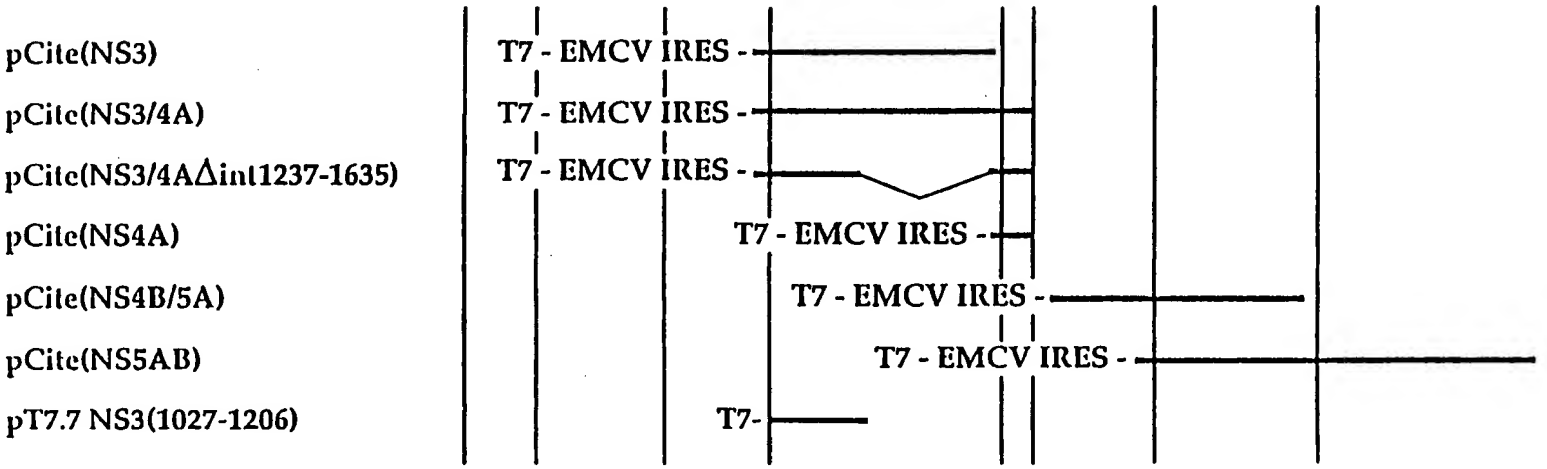
7. Use of the compositions of matter according to claims 5 and 6 to set up an enzymatic test capable of selecting, for therapeutic purposes, compounds that inhibit the enzymatic activity associated with NS3.

8. Method for reproducing in vitro the proteolytic activity of the HCV NS3 protease, compositions of matter and use of said compositions of matter to set up an

enzymatic test capable of selecting, for therapeutic purposes, compounds that inhibit the enzymatic activity associated with NS3, according to the above description, examples and claims.



Plasmid



T7 : Bacteriophage T7 promoter.

EMCV IRES: internal ribosome entry site of encephalomyocarditis virus.

INTERNATIONAL SEARCH REPORT

Interna al Application No
PCT/IT 95/00018

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 A61K38/43 A61K38/48 G01N33/48 C12Q1/37

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 A61K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	JOURNAL OF VIROLOGY, vol. 68, no. 6, 1994 pages 3753-3760, C. FAILLA ET AL. 'Both NS3 and NS4A are required for proteolytic processing of HCV nonstructural proteins' *see the whole article* ---	1-8
A	JOURNAL OF VIROLOGY, vol. 67, no. 7, 1993 pages 4017-4026, L. TOMEI ET AL. 'NS3 is a serine protease required for processing of HCV polyprotein' *see the whole article* ---	

-/--

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 July 1995	04.08.95

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Td. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Marie, A

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VIROLOGY, vol. 204, 1994 pages 163-169, S.S. LEINBACH ET AL. 'Substrate specificity of the NS3 proteinase of HCV...' *see the whole article*	
A	PROC. NATL. ACAD. SCI. USA, vol. 91, 1994 pages 888-892, E. PIZZI ET AL. 'Molecular model of the specificity pocket of the HCV protease...' *see the whole article*	